Artifacts Detection for Diffusion Tensor Imaging using Convolutional Neural Networks

Xiaofu He[a]*, Zijing Wang[b]*, Sikai Huang[c]*, Lihui Pan[d], Peilin Li[d], Emily C Merz[e], Pooja M. Desai[f], Rehan Rehman[g], Kimberly G. Noble[d] , Feng Liu

[a] Department of Psychiatry, Columbia University & New York State Psychiatric Institute, New York, NY 10032

[b] Data Science Institute, Columbia University, New York, NY 10032

[c] Department of Computer Science, Columbia University, New York, NY 10032

[d] Data Science Institute, Columbia University, New York, NY 10032

[c] Department of Statistics, Columbia University, NY 10032

[d] Teachers College, Columbia University, NY 10032

*Corresponding author: Xiaofu He, E-mail: xh2170@cumc.columbia.edu

Abstract

Diffusion Tensor Imaging (DTI) is very sensitive to subject motion. This issue is particularly problematic and can result in especially severe artifacts when attempting to collect DTI data on infants, children, and certain clinical populations that suffer from mental illnesses, for instance attention deficit hyperactivity disorder. Therefore, to salvage artifacts-corrupted DTI data, artifacts detection is important for DTI studies of these populations. Most existing artifacts detection methods are volume-based, which will exclude the whole volume where slices contain artifacts. However, the excluded volumes sometimes only contain a few slices with artifacts, i.e., the majority of the slices in that volume can be used for subsequent fitting for tensor estimation. Therefore, in order to efficiently screen out as many acceptable slides as possible and speed up efficiency, we propose a new slice-based artifacts detection method for DTI data using 2D Convolutional Neural Networks (CNNs). The proposed method is trained

and tested on a DTI dataset from 7 subjects (18□104 volumes / subject, 56□81 slices / volume, 28490 slices in total, all subjects age range between 5-9(?) years old). The F1 score of detecting slices with artifacts using CNNs is ~77%, while the score using Support Vector Machine (SVM) is only around 64%, which indicates that our method can detect artifacts in DTI data more efficiently.

Keywords: Diffusion Tensor Imaging (DTI), Deep Learning, Convolutional Neural Networks (CNNs), Quality Control (QC), TensorFlow

1. Introduction

Diffusion tensor imaging (DTI) is one of the magnetic resonance imaging (MRI) modalities that measure the water molecules diffusion patterns in biological tissue (Basser and Pierpaoli 1996, Cook, McCarthy et al. 2017). It has become a useful tool for the noninvasive study of intrinsic white matter properties, and of structural connectivity in the human brain (Hagmann, Jonasson et al. 2006, Alexander, Lee et al. 2007), which, for example, has emerged as a breakthrough for the Human Brain Connectome project (Toga, Clark et al. 2012){Toga, 2012, Mapping the human connectome}. DTI can also be used to study brain maturation from infancy to childhood and into adulthood (Hüppi, Maier et al. 1998, Mukherjee, Miller et al. 2002, Schneider, Il'yasov et al. 2004, Bashat, Sira et al. 2005, Evans and Group 2006, Huang, Zhang et al. 2006, Hüppi and Dubois 2006, Cascio, Gerig et al. 2007, Qiu, Tan et al. 2008, Faria, Zhang et al. 2010, Neuner, Kupriyanova et al. 2010, Sanchez, Richards et al. 2012). Diffusion anisotropy can be modeled by estimating a diffusion tensor at each spatial location in the brain from a series of measures derived from diffusion weighted imaging (DWI), which is acquired by applying diffusion gradients along many non-collinear orientations in a 3D space. By fitting a diffusion tensor to the observed measurements and using a linear least squares method, the principal direction (PD) of the tensor is expected to reflect the direction of the local diffusion of water molecules. Using this method, one can derive various useful diffusion anisotropy indices, such as fractional anisotropy (FA) (Basser and Pierpaoli 1996, Beaulieu 2002), which quantifies how strongly directional the local tract structure is. Also, this method is useful in localizing brain changes related to neuropsychiatric and movement disorders, including bulimia nervosa (He, Stefan et al. 2016), Depression (Yin, He et al. 2016), obsessive-compulsive disorder, (He, Steinberg et al. 2018), Tourette syndrome (Neuner, Kupriyanova et al. 2010, Liu, Miao et al. 2013) and Parkinson's disease (Meijer, Bloem et al. 2013, Prodoehl, Li et al. 2013).

However, DTI data analysis is subject to distortions related to artifacts resulting from subject motion, which will alter the molecular motion that is being measured by the MRI scanner, thereby causing signal changes in DWI (Chang, Jones et al. 2005, Liu, Wang et al. 2010, Jones 2011, Zhou, Liu et al. 2011, Lauzon, Caffo et al. 2012, Lauzon, Asman et al. 2013, He, Liu et al. 2014). These motion artifacts are particularly crucial for imaging data acquired from non-sedated populations, for instance infants and children, as well as from patients with attention deficit hyperactivity and movement disorders, such as those with Parkinson's disease and Tourette syndrome (Skare, Hedehus et al. 2000, Bastin 2001, Epstein, Casey et al. 2007, Nair and Hu 2012, Power, Barnes et al. 2012, Van Dijk, Sabuncu et al. 2012). These artifacts can result in erroneous DWI measures and further produce incorrect estimations of diffusion tensors which will affect the diffusion anisotropy indices such as FA images. Thus, the artifacts eventually affect the sensitivity and specificity of DTI-based biomarkers, and complicate the understanding of normal development and pathological processes (Pierpaoli and applications 2010).

There are usually two types of artifacts detection and correction methods: during DTI data acquisition and post-processing. The former refers to methods that propose to detect and correct artifacts of DTI data during the acquisition phase, including systematic vibration artifacts detection (Atkinson, Porter et al. 2000, Miller and Pauly 2003, Gallichan, Scholz et al. 2010, Kober, Gruetter et al. 2012, Mohammadi, Nagy et al. 2012) and external devices (Rohde, Barnett et al. 2004, Zaitsev, Dold et al. 2006, Qin, van Gelderen et al. 2009, Forman, Aksoy et al. 2010). Sewonu (Sewonu, Hossu et al. 2013) conducted an automated DTI quality control (QC) protocol based on the American College of Radiology procedures. The navigator techniques have been developed to track and correct motion (Qin, Schmidt et al. 2013, Godenschweger, Kägebein et al. 2016). Post-processing detection of artifacts refers to artifacts detection

after the data collection process has been completed, and various post-processing methods have been proposed to address this type of artifacts detection. Mangin (Mangin, Poupon et al. 2002) proposed a new tensor estimation method which uses maximization of mutual information from the Geman-McLure M-estimator during the procedure in order to reduce artifacts in DWI data. Rohde (Rohde, Barnett et al. 2004) implemented a registration technique and a spatial transformation model to design a method for correcting DWI volumes spatial misalignment. Chang (Chang, Jones et al. 2005) designed a tensor estimation approach named robust estimation of tensors by outlier rejection using the iteratively reweighted least-squares regression method. Ding (Ding, Gore et al. 2005) presented a novel smoothing technique based on the traditional method of filtering anisotropic diffusion images to reduce noise in DTI data by allowing isotropic smoothing within homogeneous regions and anisotropic smoothing along structure boundaries. Dubois (Dubois and Girard 2006) proposed a method for generating schemes of diffusion gradient orientations that allow the diffusion tensor to be reconstructed from partial datasets in clinical DTI data using a general energy-minimization electrostatic model, in which the interactions between orientations are weighted according to their temporal order during acquisition. Jiang (Jiang, Van Zijl et al. 2006) developed a DTI toolbox named DtiStudio, which provides manually and visually inspection of the DTI data. Hasan (Hasan 2007) presented a theoretical framework and a parameter optimization method for DTI QC on the basis of the analytical error propagation of the mean diffusivity. Jiang (Jiang, Chou et al. 2009) proposed an outlier detection and exclusion method for DTI data through testing the consistency in the apparent diffusion coefficient (ADC), which is a detection criterion based on the smoothness of the fitted surface of the peanut-shaped ADC directional profile. Liu (Liu, Wang et al. 2010) designed an automatic DWI QC framework and developed a pipeline named DTIPrep, which uses normalized correlation between

successive slices and across all the diffusion gradients to screen artifacts. Oguz (Oguz, Farzinfar et al. 2014) proposed an improved version of the DTIPrep software with functions of detecting and removing the directional artifacts and applying bias analysis. Zhou (Zhou, Liu et al. 2011) presented an automatic method for identifying and rejecting outlier images using local binary patterns and 2D partial least square to robustly estimate diffusion tensors. Liu (Liu, Liu et al. 2012) proposed an improved method on the correction of eddy current-induced distortion in DWI images through excluding signals from the cerebral spinal fluid. Lauzon (Lauzon, Caffo et al. 2012, Lauzon, Asman et al. 2013) designed a parallel pipeline to process quantitative DTI QC through combing wild bootstrap with simulation extrapolation, and to assess the quality of DTI data based on FA power calculation. Farzinfar (Farzinfar, Li et al. 2013) presented a tensor evaluation method with a consistent threshold definition using the Monte-Carlo simulation, which judges DWI data according to two error metrics based on directional distribution bias of FA and PD. Farzinfar (Farzinfar, Oguz et al. 2013) also proposed an entropy-based QC method, which could be used as an indication of clusters in the brain image. Andersson (Andersson and Sotiropoulos 2015) designed a non-parametric framework to detect the signal dropout and slices affected by the spurious signal loss, and{Andersson, 2016, Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images} also proposed a slice-to-volume model based on the signal loss to detect and correct the intra-volume movement (Andersson, Graham et al. 2016, Andersson, Graham et al. 2017). Validation performed on the Whitehall Imaging sub-study data (Filippini, Zsoldos et al. 2014) and simulated data indicated that this model, which was implemented in the FMRIB Software Library (FSL) EDDY, performed better than the previous volume-to-volume model. Bastiani (Bastiani, Cottaar et al. 2018) proposed two automated QC tools, which are highly dependent on FSL EDDY, i.e.,

QUality Assessment for DMRI and Study-wise QUality Assessment for DMRI, to apply on single subject QC and group QC respectively. However, FSL EDDY is computationally expensive and deficient in flexibility of parameter adjustment for false-positive cases that fail running artifacts detection and correction process.

Thus, it is possible to improve these DTI QC methods further. However, volume-based DTI QC methods usually exclude volumes that contain even a few bad slices, which results in wasting a great number of usable DWI images. Besides, there is a growing body of research focusing on reconstructing the whole volume based on slices. Therefore, in this paper, we propose a new method using slice instead of volume as a detection basis. It takes advantage of all volumes to screen out bad slices in different DTI directions and reserve most usable volumes, which would provide the foundation for further reconstruction-related researches. In order to enhance the efficiency of judgement, we combined the slice-based idea with Convolutional Neural Networks (CNNs) which is an efficient tool in pattern recognition field (LeCun, Bottou et al. 1998). Kelly (Kelly, Pietsch et al. 2016) tried to apply CNNs and decision trees to train and test on manually labelled volumes. Since manually labelling volumes is time-consuming, Graham (Graham, Drobnjak et al. 2018) proposed an approach that trains on simulated data and tests on a small volume of manually labelled real data using CNNs. Samani(Samani, Alappatt et al. 2020) adopted CNN model and transfer learning aiming at detecting different kinds of artifacts including motion. Ettehadi (Ettehadi, Zhang et al. 2021) developed a two-step CNN-based model to discriminate the quality of the whole 3D DTI image.

2. Methods

In this paper, we build CNNs using TensorFlow to classify and evaluate quality of DTI.

2.1 Participants

Participants were selected from our socioeconomic study, which recruited subjects from community events and posting flyers in local neighborhoods in New York. A socioeconomically diverse sample was recruited by ensuring that families in the study represented a wide range of parental educational attainment. Interested families were contacted by phone and screened for eligibility. Inclusionary criteria for children were the following: 1) between 5 and 9 years of age, 2) born after 37 weeks of gestation, 3) born from a singleton pregnancy, 4) no history of medical or psychiatric problems, 5) the primary caregiver and child were fluent in English which is the language spoken most often in the home. Children with contraindications for MRI scanning were excluded.

Children ranged from 5.12 to 7.84 years of age (60% female), family income ranged from $8,976 to $300,000, and parental education ranged from 11.50 to 20.00 years. The racial and ethnic composition of child participants was 50% Hispanic/Latino, 20% African American, non-Hispanic/Latino, and 25% White, non-Hispanic/Latino. Full sample characteristics are provided in Table 1.

Table 1

2.2 Data acquisition

DTI data were acquired on a General Electric (GE) MR750 3.0 Tesla whole body scanner (GE Medical Systems, Waukesha, Wisconsin). Twenty children participated in the DTI data acquisition procedure. Before the DTI data acquisition session, children were trained to keep still during their lying in a mock scanner. We used a single-shot spin echo planar imaging sequence along 15 non-collinear spatial directions to collect the diffusion data (TR = 15700 ms, TE = minimum, flip angle = 90, field of view = 24 cm, matrix size = 132 × 128 and machine-interpolated to 256 × 256 for post-processing, b value = 1000 s/mm2, and three baseline images at b = 0 s/mm2 , slice thickness = 2.5 mm without gap, an in-plane resolution of 0.94 mm, and 60 axial slices).

The scan duration of each excitation was 4 minutes and 43 seconds, and the average of each participant's two separate excitations (Number of Excitation = 2) was kept.

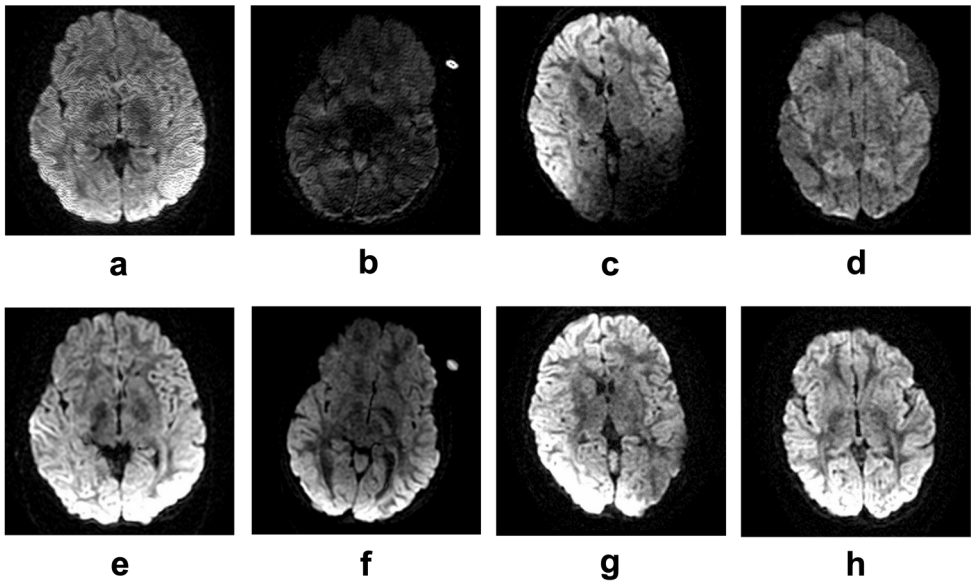

Figure 1

2.3 Proposed Model

CNN is a special neural network inspired by biological connectivity between neurons of the human brain, which consists of a multiple specific layers designed for pattern recognition from images (LeCun, Bottou et al. 1998, LeCun 2015). This method has been successfully applied into practical visual images analysis (Ciresan, Meier et al. 2011, Cireşan, Meier et al. 2012). Therefore, we decided to use this method in our DTI QC experiment and designed CNNs to train and test our model based on DWI images (Fig. 1) which were manually labeled by P. M. Desai and R. Rehman.

CNN typically includes convolutional, pooling and fully connected layers. Convolutional layer applies convolution operation on the raw image and outputs a feature map. Pooling layer is designed to reduce the dimensionality of the feature map to decrease processing time. Fully connected layer contains neurons that have connections with all neurons in the previous layer and is used to compute the class

score using Softmax activation function. (Le Callet, Viard-Gaudin et al. 2006, Krizhevsky, Sutskever et al. 2012)

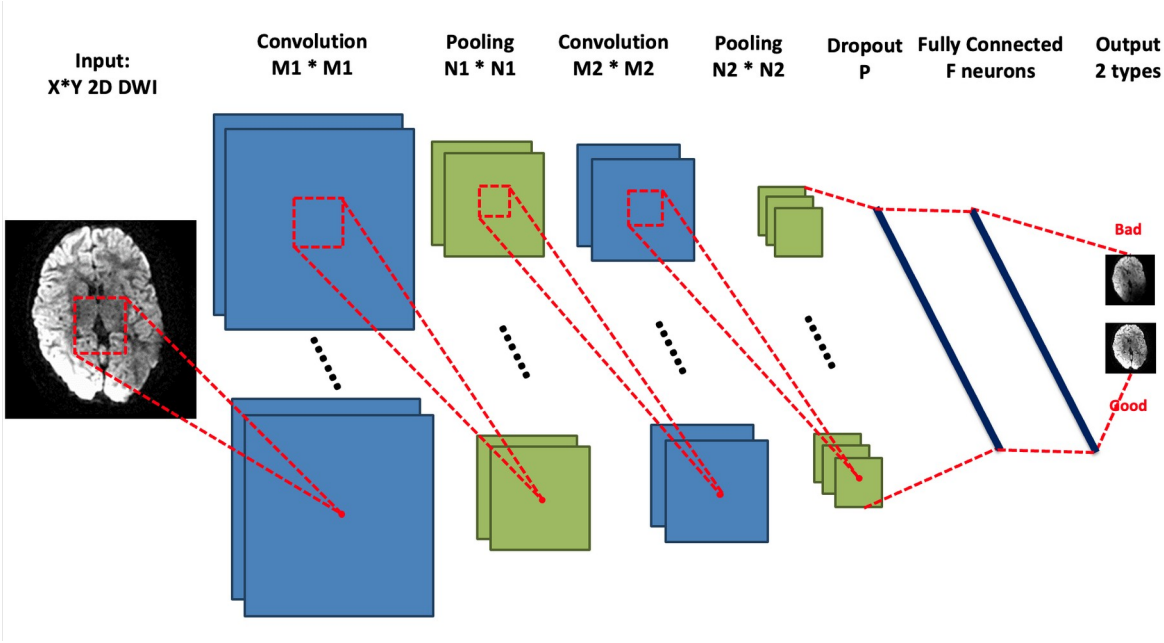

Figure 2

The overall structure of our CNNs model contains 8 layers with weights and bias (Fig. 2), two convolutional layers and two pooling layers are mutually crosswise arranged, a dropout layer, a flatten layer, a fully connected layer, and an output layer. The output of the fully connected layer is fed to the output layer with a Softmax activation function, the result of which indicates the quality of DWI images as good or bad.

The input of our model is a 2D DWI image with size of $X \times X$. Firstly, the image is filtered by the first convolutional layer with $K1$ kernels of size $M1 \times M1$. Here we use ReLu (Nair and Hinton 2010) as the activation function of the convolutional layer. Then the output from first layer is pooled in the first pooling layer with the down-sampling window of size $N1 \times N1$ and a stride of size $P1$. After that, the pooled data is filtered by the second convolutional layer with $K2$ kernels of size $M2 \times M2$, the activation function is also ReLu as in previous convolutional layer. Then the output is pooled again in the second pooling layer with the down-sampling window of size $N2 \times N2$ and a stride of

size $P2$. After applying a dropout layer with probability of P, we flattened the activation map. Finally, the fully connected layer with F neurons calculates the class score and the output layer outputs the final result.

2.3.1 Reducing overfitting

It is very common that data is insufficient because of the large number of parameters and hyper-parameters to be trained. This problem is exacerbated with by deep neural networks with millions of parameters since it leads to overfitting in models. To prevent overfitting, we used three methods: Dropout, L2 norm regularization and early stopping.

2.3.1.1 Dropout

Dropout sets the output of each hidden neuron to zero with a certain probability and the dropped-out neurons don't contribute to the training process, which prevents a hidden unit from relying on other hidden units (Srivastava, Hinton et al. 2014). Meanwhile, dropout is an efficient method of performing model averaging with a great number of neural networks to reduce the test set error since randomly dropping out hidden units could train various separate neural networks in a reasonable time (Hinton, Srivastava et al. 2012). However, a drawback of the dropout method is that the testing data must be kept to the same expectation with training data. Traditional dropout needs to scale test data while inverted dropout only scales train data, which would save much time during the testing process. Since we plan to use our final model to design a system to do QC automatically, in order to reduce the testing time, we chose the inverted dropout method.

2.3.1.2 L2 norm regularization

L2 norm regularization (Ng 2004) adds a weight decay term to prevent overfitting, which modifies the learning rule to multiplicatively shrink the weight vector by a

constant factor. Since our network structure is not very deep, we applied L2 norm regularization during the optimization. The cost function is defined as below:

$$w^{*} = \arg\min_{w} \left\{ \frac{1}{N} \sum_{i=1}^{N} Loss_i + \underbrace{\rho \sum_{l=1}^{L} \sum_{m=1}^{M_l} w_{l,m}^2}_{weight\,decay\,term} \right\} \quad (1)$$

where ρ is a constant that controls the regularization, N is the size of the mini-batch that contains DWI images inputted into the CNNs model, $Loss_i$ is the loss of i^{th} DWI image in the mini-batch, L represents the number of layers and M_l is the number of weights in layer l. L2 regularization is computational efficient since it has analytical solutions in the backward of training process and it tends to induce weights to be small value, which may release the large weights on certain units to other units with smaller weights.

2.3.1.3 Early stopping

Early stopping (Yao, Rosasco et al. 2007) solves the overfitting problem by interrupting the training procedure once the accuracy of predicting on validation set goes beyond a certain bound. Early stopping (Yao, Rosasco et al. 2007) solves the overfitting problem by interrupting the training procedure once the performance of the prediction on validation set has achieved a certain bound. In this paper we monitored the prediction loss on the validation set and the early stopping would be triggered if there's no improvement of the training performance after several epochs. After every epoch, we evaluated the prediction loss on the validation set, and save the model when the error on the validation set does not decrease over a given number of epochs such as 1000 epochs. When training process finished, we picked an optimal model with highest τ from the model pool, where:

$$\tau = \frac{1}{\frac{1}{true_{positive}} + \frac{1}{true_{negative}}}$$

2.3.2 Optimization

2.3.2.1 Softmax function and Cross-entropy

Softmax (Bishop 2006) function is often used as the activation function of the output layer for multiclass classification, which is defined as below:

$$q(y=h|x, W, b) = softmax(f_h(x; w, b)) = \frac{e^{f_h}}{\sum_{j=1}^{K} e^{f_j}} \quad 2$$

where x represents the original images input into the model, h is the ground truth label of x, K is the number of classes, W and b are the parameters in the output layer.

Cross-entropy (De Boer, Kroese et al. 2005) is the most commonly used loss function in neural networks. It works well when the hypothesis of outputs to neural networks is independent, and activations can represent the probability that each hypothesis is true. Intuitionally, cross-entropy calculates the distance between true distribution and estimated distribution, which is defined as below:

$$H(p, q) = -\sum_{x} p(x) \log q(x) \quad 3$$

where p is the ground-truth posterior and q is the estimated posterior.

For the convenience, assume that data x has ground truth label y_i, we combined Softmax and cross-entropy formulas above:

$$p(y=y_i|x, W, b) = 1 \quad 4$$

$$p(y \neq y_i|x, W, b) = 0 \quad 5$$

Then, the cross-entropy loss is calculated as below:

$$Loss_i = -\sum_{j=1}^{K} p(y=j|x, W, b) \log q(y=j|x, W, b) \quad (6)$$

$$¿ -\log q(y=y_i|x, W, b) \quad (7)$$

$$¿ -\log \frac{e^{f_{y_i}}}{\sum_{j=1}^{K} e^{f_j}} \quad (8)$$

Optimization process in training is to find the optimal solution to the cost function, which is defined as:

$$Cost = \frac{1}{N} \sum_{i=1}^{N} Loss_i \quad (9)$$

2.3.2.2 Adaptive Moment Estimation (Adam)

To reduce the convergence time and improve the performance of our CNNs model, we chose Adaptive Moment Estimation (Adam) as the optimizer, which is an updated version of Root Mean Square Propagation (RMSProp) (Kingma and Ba 2014). Adam makes use of both Momentum and the RMSProp so it can handle sparse gradients under non-stationary settings (Duchi, Hazan et al. 2011, Tieleman and Hinton 2012). The process is defined as below:

Momentum: $$V_{dw} = \beta_1 V_{dw} + (1-\beta_1) dw,\ V_{db} = \beta_1 V_{db} + (1-\beta_1) db \quad (10)$$

RMSProp: $$S_{dw} = \beta_2 S_{dw} + (1-\beta_2) dw^2,\ S_{db} = \beta_2 S_{db} + (1-\beta_2) db^2 \quad (11)$$

Bias correction:

$$\begin{cases} V_{dw}^{¿} = \dfrac{V_{dw}}{(1-\beta_1^t)} \\ S_{dw}^{¿} = \dfrac{S_{dw}}{(1-\beta_2^t)} \end{cases} \quad (12)$$

$$\begin{cases} \hat{V}_{db} = \dfrac{V_{db}}{(1-\beta_1^t)} \\ \hat{S}_{db} = \dfrac{S_{db}}{(1-\beta_2^t)} \end{cases} \quad 13$$

Update:

$$\begin{cases} w = w - \alpha \dfrac{\hat{V}_{dw}}{\sqrt{\hat{S}_{dw}} + \varepsilon} \\ b = b - \alpha \dfrac{\hat{V}_{db}}{\sqrt{\hat{S}_{db}} + \varepsilon} \end{cases} \quad 14$$

where α is the step size, β_1 and β_2 are exponential decay rates for the moment estimates, ε is used to prevent any division by zero, and t is the time step. We used the default values of α=0.0001, β_1=0.9, β_2=0.999, ε=1e-8, t=0 from the original paper (Kingma and Ba 2014).

2.3.2.3 Learning rate

Training is a time-consuming process and we implemented learning rate decay to speed up the learning step. As the learning step iterates, the cost will tend towards the minimum value, but won't converge because of the fixed learning rate α. Therefore, if the learning rate reduces slowly during training, the step size will be smaller where the cost is close to the optimal value, and the optimizer will end up oscillating in a tighter region near the optimal solution (Darken, Chang et al. 1992). The learning rate decay updating method is defined as below:

$$\alpha = \frac{\alpha_0}{1 + decay_{rate} * iterate_{num}} \quad 15$$

where α_0 is the initial learning rate, $iterate_{num}$ is the iteration number of training, and $decay_{rate}$ is the decay rate used to reduce the learning rate (see Table S1).

2.3.3 Imbalanced data

Since our dataset is not large enough, the performance of model is highly affected by the imbalanced data. Currently there are ~ 90% samples are good and ~10% samples are bad in our dataset. Therefore, we tried data augmentation methods and add class weight parameter to the CNN model to balance the class samples and solve this issue.

Data augmentation is the most common and simple way to increase the number of samples in the dataset and can also reduce overfitting problem within deep learning model due to the imbalanced data (Krizhevsky et al. (2012)). Dan et al. employ several geometric transformations methods like rotation, scaling and translation to increase the samples (Dan et al. (2012)). Suppose we have n_g instances of good quality, n_b instances of bad quality in our training set, respectively ($N_{total\,training\,samples} = n_g + n_b$).

In our preprocessing steps, we utilized the geometric transformation methods like rotation (up to ± 5 °) to increase the number of samples in the minority class (bad class). We then defined r as the ratio of bad samples in the training set after the rotation (see Table S1) and $n_b^{'}$ as number of bad images we should generated through the rotation based on the calculation under different ratio r.

We have tried two approaches to randomly select the bad samples from training set, which would be used to generate the augmented samples. For the first approach, we randomly picked $\frac{n_b^{'}}{10}$, marked as *s*samples (10 rotation degrees up to ± 5 °) from all of the bad samples in the training set (represented by n_b). In another word, for every sample in the *s*, we would generate 10 rotated images and add them to our training set.

As the second approach, we have 2 steps within that. For the step 1, we took $\frac{n_b^{'}}{n_b}$ as d, which represents number of new images that every bad sample in the training set should generate. For the augmentation, we randomly selected d degrees (without duplicate) from 10 rotation degrees and then for every bad sample in n_b, generating in total d rotated images regarding every select degree. We already got $d*n_b$augmented images in total after step 1. Now we will move to the step 2. In most cases, $n_b^{'}-d*n_b>0$, so for the remaining samples to be augmented, what we call d_1 ($d_1=, n_b^{'}-d*n_b$), we randomly picked d_1 samples from n_b. After that, for every sample in d_1, we would rotate it by a randomly selected degree from 10-d, which is the remaining degrees from step 1. In this way we can ensure that there's no duplicate sample (same original image and same rotation degree) generated from the data augmentation step.

3. Results

After tuning both hyper-parameters and other parameters in practical experiments, we obtained optimal values for our model (see Table S1 and Table S2). For several hyper/parameters that previous researches have given widely accepted values or don't really influence the process much, default values could be optimal values. For example, exponential decay rates for moment estimates β_1 and β_2, and division term ε.

After training, we found a good CNNs structure with 2 convolutional layers, 2 pooling layers, 1 fully connected layer and 1 output layer. There are 17 filters of size 7 $\times$ 7 with stride of 1 in the 1st convolutional layer and 7 filters of size 5 $\times$ 5 also with stride 1 in the 2nd convolutional layer. The neuron number of fully connected layer is 100 and the dropout rate is 0.4, and our model converge at 15000 iterations.

After tuning on the validation set, Table S3 shows that under our best CNNs model with learning rate = 0.001 and the bad slice ratio r=0.4, the F1-score of is 77%.In comparison, we also applied Support Vector Machine (SVM) method, which is a classical method for two-group classification problem (Cortes and Vapnik 1995) and has proven performance in image classification problems, on the same data set. More specifically, in our SVM model, the regularization parameter is 1.0 using squared L2 penalty and kernel type is radial basis function. The final F1-score is 64% on the test data set, which is about 13% lower than our newly-put-forward model.

4. Discussion

In this paper, we presented a 2D CNNs-based DTI QC method, which extracted artifacts features using fewer layers and with greatly reduced numbers of parameters than traditional neural networks by using convolutional kernels. As such, we have demonstrated that our model is more efficient compared with traditional methods such as SVM. Along with that, we also tried different architecture and parameters to present a stable model with high performance. For example, to avoid overfitting, we not only used dropout layer as shown in the final model but also added batch normalization layers at the very beginning, which did not address better performances, thus indicated more research in the future. What's more, instead of doing quality check volume by volume, our approach concentrates on the slice-level detection and exclusion. In this way, it would be able to keep as much gradient direction as possible and provide more comprehensive information for the subsequent tensor reconstruction.

There are also some limitations to the method proposed in our work. The dataset used in this paper is relatively small and manual labeling is time-consuming. Graham (Graham, Drobnjak et al. 2018) mentioned that simulated data could be a great supplement to training and test datasets, which may produce a more balanced dataset

that helps improve the CNNs model. Also, we can test our model on large publicly available datasets although manually labelling is challenging. We can also compare the Factional anisotropy (FA) measurements between our approach and other volume-based approaches under large publicly available datasets to further quantify our improvement in the post-analysis. Moreover, in future works, we could try other deep learning methods like fully CNNs (Dolz, Desrosiers et al. 2017), which simply replaces FC layer with 1*1 size kernel to make computation more efficient or save memories. Moreover, U-Net (Ronneberger, Fischer et al. 2015) uses limited data more efficiently by using data augmentation strategy, we suppose this model could detect the artifacts patterns in bad slices. Besides, we could combine traditional deep learning methods with some Bayesian methods (Garnelo, Rosenbaum et al. 2018). For example, the neural network is a discriminative model which has a good performance on function approximation, Gaussian process is one of the Bayesian methods and could infer a function at test time based on the prior. Combining the advantages of both methods will make our model train more accurate and efficient.

Acknowledgments

This study was supported in part by a NARSAD Young Investigator Grant (PI: X. He) from the Brain & Behavior Research Foundation, The New York State Psychiatric Institute MRI Pilot Award (PI: X. He), Columbia Irving Institute Imaging Pilot Award (PI: X. He) from NIH UL1 TR000040. We also thank Mr., and Ms. Rachel xx for the initially manual labeling DWI images.

References

Alexander, A. L., et al. (2007). "Diffusion tensor imaging of the brain." **4**(3): 316-329.

Andersson, J. L. and S. N. Sotiropoulos (2015). "Non-parametric representation and prediction of single- and multi-shell diffusion-weighted MRI data using Gaussian processes." Neuroimage **122**: 166-176.

Andersson, J. L. R., et al. (2017). "Towards a comprehensive framework for movement and distortion correction of diffusion MR images: Within volume movement." Neuroimage **152**: 450-466.

Andersson, J. L. R., et al. (2016). "Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images." Neuroimage **141**: 556-572.

Atkinson, D., et al. (2000). "Sampling and reconstruction effects due to motion in diffusion-weighted interleaved echo planar imaging." Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **44**(1): 101-109.

Bashat, D. B., et al. (2005). "Normal white matter development from infancy to adulthood: comparing diffusion tensor and high b value diffusion weighted MR images." Journal of magnetic resonance imaging **21**(5): 503-511.

Basser, P. J. and C. Pierpaoli (1996). "Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI." J Magn Reson B **111**(3): 209-219.

Bastiani, M., et al. (2018). "Automated quality control for within and between studies diffusion MRI data using a non-parametric framework for movement and distortion correction." Neuroimage.

Bastin, M. E. (2001). "On the use of the FLAIR technique to improve the correction of eddy current induced artefacts in MR diffusion tensor imaging." Magn Reson Imaging **19**(7): 937-950.

Beaulieu, C. (2002). "The basis of anisotropic water diffusion in the nervous system–a technical review." NMR in Biomedicine **15**(7-8): 435-455.

Bishop, C. M. (2006). Pattern Recognition and Machine Learning, Springer.

Cascio, C. J., et al. (2007). "Diffusion tensor imaging: application to the study of the developing brain." Journal of the American Academy of Child & Adolescent Psychiatry **46**(2): 213-223.

Chang, L. C., et al. (2005). "RESTORE: robust estimation of tensors by outlier rejection." Magn Reson Med **53**(5): 1088-1095.

Cireşan, D., et al. (2012). "Multi-column deep neural networks for image classification."

Ciresan, D. C., et al. (2011). Flexible, high performance convolutional neural networks for image classification. IJCAI Proceedings-International Joint Conference on Artificial Intelligence, Barcelona, Spain.

Cook, J. M., et al. (2017). "Older Adults with PTSD: Brief State of Research and Evidence-Based Psychotherapy Case Illustration." The American Journal of Geriatric Psychiatry **25**(5): 522-530.

Cortes, C. and V. Vapnik (1995). "Support-vector networks." Machine learning **20**(3): 273-297.

Darken, C., et al. (1992). Learning rate schedules for faster stochastic gradient search. Neural Networks for Signal Processing [1992] II., Proceedings of the 1992 IEEE-SP Workshop, IEEE.

De Boer, P.-T., et al. (2005). "A tutorial on the cross-entropy method." **134**(1): 19-67.

Ding, Z., et al. (2005). "Reduction of noise in diffusion tensor images using anisotropic smoothing." Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **53**(2): 485-490.

Dolz, J., et al. (2017). "3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study." Neuroimage.

Dubois, L. and M. Girard (2006). "Early determinants of overweight at 4.5 years in a population-based longitudinal study." International journal of obesity **30**(4): 610.

Duchi, J., et al. (2011). "Adaptive subgradient methods for online learning and stochastic optimization." **12**(Jul): 2121-2159.

Epstein, J. N., et al. (2007). "Assessment and prevention of head motion during imaging of patients with attention deficit hyperactivity disorder." Psychiatry Research: Neuroimaging **155**(1): 75-82.

Ettehadi, N., et al. (2021). Automatic Volumetric Quality Assessment of Diffusion MR Images via Convolutional Neural Network Classifiers. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC).

Evans, A. C. and B. D. C. Group (2006). "The NIH MRI study of normal brain development." Neuroimage **30**(1): 184-202.

Faria, A. V., et al. (2010). "Atlas-based analysis of neurodevelopment from infancy to adulthood using diffusion tensor imaging and applications for automated abnormality detection." Neuroimage **52**(2): 415-428.

Farzinfar, M., et al. (2013). "DTI Quality Control Assessment via Error Estimation From Monte Carlo Simulations." Proc SPIE Int Soc Opt Eng **8669**: 1667549.

Farzinfar, M., et al. (2013). "Diffusion imaging quality control via entropy of principal direction distribution." Neuroimage **82**: 1-12.

Filippini, N., et al. (2014). "Study protocol: the Whitehall II imaging sub-study." BMC psychiatry **14**(1): 159.

Forman, C., et al. (2010). Self-encoded marker for optical prospective head motion correction in MRI. International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer.

Gallichan, D., et al. (2010). "Addressing a systematic vibration artifact in diffusion-weighted MRI." Human brain mapping **31**(2): 193-202.

Garnelo, M., et al. (2018). "Conditional Neural Processes." arXiv preprint arXiv:1807.01613.

Godenschweger, F., et al. (2016). "Motion correction in MRI of the brain." Physics in Medicine & Biology **61**(5): R32.

Graham, M. S., et al. (2018). "A supervised learning approach for diffusion MRI quality control with minimal training data." Neuroimage **178**: 668-676.

Hagmann, P., et al. (2006). "Understanding diffusion MR imaging techniques: from scalar diffusion-weighted imaging to diffusion tensor imaging and beyond." Radiographics **26**(suppl_1): S205-S223.

Hasan, K. M. (2007). "A framework for quality control and parameter optimization in diffusion tensor imaging: theoretical analysis and validation." Magn Reson Imaging **25**(8): 1196-1202.

He, X., et al. (2014). "Automated assessment of the quality of diffusion tensor imaging data using color cast of color-encoded fractional anisotropy images." Magn Reson Imaging **32**(5): 446-456.

He, X., et al. (2016). "Altered White Matter Microstructure in Adolescents and Adults with Bulimia Nervosa." Neuropsychopharmacology **41**(7): 1841-1848.

He, X., et al. (2018). "Altered frontal interhemispheric and fronto-limbic structural connectivity in unmedicated adults with obsessive-compulsive disorder." Hum Brain Mapp **39**(2): 803-810.

Hinton, G. E., et al. (2012). "Improving neural networks by preventing co-adaptation of feature detectors."

Huang, H., et al. (2006). "White and gray matter development in human fetal, newborn and pediatric brains." Neuroimage **33**(1): 27-38.

Hüppi, P. S. and J. Dubois (2006). Diffusion tensor imaging of brain development. Seminars in Fetal and Neonatal Medicine, Elsevier.

Hüppi, P. S., et al. (1998). "Microstructural development of human newborn cerebral white matter assessed in vivo by diffusion tensor magnetic resonance imaging." Pediatric research **44**(4): 584.

Jiang, H., et al. (2009). Outlier detection for diffusion tensor imaging by testing for ADC consistency. Proceedings of 17th Annual Meeting ISMRM. San Diego, California, USA.

Jiang, H., et al. (2006). "DtiStudio: resource program for diffusion tensor computation and fiber bundle tracking." Computer methods and programs in biomedicine **81**(2): 106-116.

Jones, D. K. (2011). Diffusion MRI: Theory, Methods, and Applications, OUP USA.

Kelly, C., et al. (2016). Transfer learning and convolutional neural net fusion for motion artefact detection. Proc. Intl. Soc. Mag. Reson. Med.

Kingma, D. P. and J. J. a. p. a. Ba (2014). "Adam: A method for stochastic optimization."

Kober, T., et al. (2012). "Prospective and retrospective motion correction in diffusion magnetic resonance imaging of the human brain." Neuroimage **59**(1): 389-398.

Krizhevsky, A., et al. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems.

Lauzon, C. B., et al. (2013). "Simultaneous analysis and quality assurance for diffusion tensor imaging." PloS one **8**(4): e61737.

Lauzon, C. B., et al. (2012). Towards automatic quantitative quality control for MRI. Medical Imaging 2012: Image Processing, International Society for Optics and Photonics.

Le Callet, P., et al. (2006). "A convolutional neural network approach for objective video quality assessment." IEEE Trans Neural Netw **17**(5): 1316-1327.

LeCun, Y. (2015). "LeNet-5, convolutional neural networks." URL: http://yann.lecun.com/exdb/lenet/.

LeCun, Y., et al. (1998). "Gradient-based learning applied to document recognition." Proceedings of the IEEE **86**(11): 2278-2324.

Liu, W., et al. (2012). Spatial normalization of diffusion tensor images with voxel-wise reconstruction of the diffusion gradient direction. International Workshop on Multimodal Brain Image Analysis, Springer.

Liu, Y., et al. (2013). "Structural abnormalities in early Tourette syndrome children: a combined voxel-based morphometry and tract-based spatial statistics study." PloS one **8**(9): e76105.

Liu, W., et al. (2010). "Quality Control of Diffusion Weighted Images." Proc SPIE Int Soc Opt Eng **7628**.

Mangin, J. F., et al. (2002). "Distortion correction and robust tensor estimation for MR diffusion imaging." Med Image Anal **6**(3): 191-198.

Meijer, F. J., et al. (2013). "Update on diffusion MRI in Parkinson's disease and atypical parkinsonism." Journal of the neurological sciences **332**(1-2): 21-29.

Miller, K. L. and J. M. Pauly (2003). "Nonlinear phase correction for navigated diffusion imaging." Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **50**(2): 343-353.

Mohammadi, S., et al. (2012). "Correction of vibration artifacts in DTI using phase-encoding reversal (COVIPER)." Magn Reson Med **68**(3): 882-889.

Mukherjee, P., et al. (2002). "Diffusion-tensor MR imaging of gray and white matter development during normal human brain maturation." American Journal of Neuroradiology **23**(9): 1445-1456.

Nair, G. and X. P. Hu (2012). "Manifestation and post hoc correction of gradient cross-term artifacts in DTI." Magn Reson Imaging **30**(6): 764-773.

Nair, V. and G. E. Hinton (2010). Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th international conference on machine learning (ICML-10).

Neuner, I., et al. (2010). "White-matter abnormalities in Tourette syndrome extend beyond motor pathways." Neuroimage **51**(3): 1184-1193.

Ng, A. Y. (2004). Feature selection, L 1 vs. L 2 regularization, and rotational invariance. Proceedings of the twenty-first international conference on Machine learning, ACM.

Oguz, I., et al. (2014). "DTIPrep: quality control of diffusion-weighted images." Front Neuroinform **8**: 4.

Pierpaoli, C. J. D. M. T., methods, and applications (2010). "Artifacts in diffusion MRI." 303-317.

Power, J. D., et al. (2012). "Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion." Neuroimage **59**(3): 2142-2154.

Prodoehl, J., et al. (2013). "Diffusion tensor imaging of Parkinson's disease, atypical parkinsonism, and essential tremor." Movement Disorders **28**(13): 1816-1822.

Qin, L., et al. (2013). "Prospective motion correction using tracking coils." Magnetic resonance in medicine **69**(3): 749-759.

Qin, L., et al. (2009). "Prospective head-movement correction for high-resolution MRI using an in-bore optical tracking system." Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **62**(4): 924-934.

Qiu, D., et al. (2008). "Diffusion tensor imaging of normal white matter maturation from late childhood to young adulthood: voxel-wise evaluation of mean diffusivity, fractional anisotropy, radial and axial diffusivities, and correlation with reading development." Neuroimage **41**(2): 223-232.

Rohde, G. K., et al. (2004). "Comprehensive approach for correction of motion and distortion in diffusion-weighted MRI." Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **51**(1): 103-114.

Ronneberger, O., et al. (2015). U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention, Springer.

Samani, Z. R., et al. (2020). "QC-Automator: Deep Learning-Based Automated Quality Control for Diffusion MR Images." Frontiers in Neuroscience **13**.

Sanchez, C. E., et al. (2012). "Neurodevelopmental MRI brain templates for children from 2 weeks to 4 years of age." Developmental Psychobiology **54**(1): 77-91.

Schneider, J. F., et al. (2004). "Fast quantitative diffusion-tensor imaging of cerebral white matter from the neonatal period to adolescence." Neuroradiology **46**(4): 258-266.

Sewonu, A., et al. (2013). "An automatic MRI quality control procedure: Multisite reports for slice thickness and geometric accuracy." **34**(4-5): 300-305.

Skare, S., et al. (2000). "Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with MRI." J Magn Reson **147**(2): 340-352.

Srivastava, N., et al. (2014). "Dropout: a simple way to prevent neural networks from overfitting." The Journal of Machine Learning Research **15**(1): 1929-1958.

Tieleman, T. and G. Hinton (2012). "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude." COURSERA: Neural networks for machine learning **4**(2): 26-31.

Toga, A. W., et al. (2012). "Mapping the human connectome." Neurosurgery **71**(1): 1-5.

Van Dijk, K. R., et al. (2012). "The influence of head motion on intrinsic functional connectivity MRI." Neuroimage **59**(1): 431-438.

Yao, Y., et al. (2007). "On early stopping in gradient descent learning." **26**(2): 289-315.

Yin, Y., et al. (2016). "Structural and Functional Connectivity of Default Mode Network underlying the Cognitive Impairment in Late-onset Depression." Sci Rep **6**: 37617.

Zaitsev, M., et al. (2006). "Magnetic resonance imaging of freely moving objects: prospective real-time motion correction using an external optical motion tracking system." Neuroimage **31**(3): 1038-1050.

Zhou, Z., et al. (2011). "Automated artifact detection and removal for improved tensor estimation in motion-corrupted DTI data sets using the combination of local binary patterns and 2D partial least squares." Magn Reson Imaging **29**(2): 230-242.

Figure legends

Table 1. Descriptive statistics for sample characteristics (N=20)

	M	SD
Child Age (Years)	6.59	0.81
Parental Education (Years)	15.13	2.52
Family Income (U.S. dollars)	80,408.00	84,122.03
	%	**n**
Child Sex		
Male	40.00	8
Female	60.00	12
Child Race/Ethnicity		
African American, non-Hispanic/Latino	20.00	4
Hispanic/Latino	50.00	10
White, non-Hispanic/Latino	25.00	5

Other	5.00	1